



日本国特許庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日      2003年10月23日  
Date of Application:

出願番号      特願2003-362750  
Application Number:

[ST. 10/C] :      [JP 2003-362750]

出願人      株式会社日立製作所  
Applicant(s):

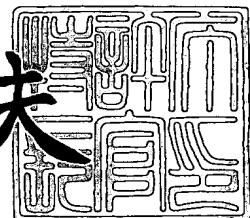
U.S. Appln. Filed 2-19-04  
Inventor: K. Shimada et al  
matthingly Stanger & Malur  
Docket H-1131

日本  
特許  
庁  
長官  
印

特許庁長官  
Commissioner,  
Japan Patent Office

今井康夫

2003年11月26日



【書類名】 特許願  
【整理番号】 K03011861A  
【あて先】 特許庁長官殿  
【国際特許分類】 G06F 12/00  
【発明者】  
【住所又は居所】 神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所  
【氏名】 島田 健太郎  
【発明者】  
【住所又は居所】 神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所  
【氏名】 橋本 顯義  
【特許出願人】  
【識別番号】 000005108  
【氏名又は名称】 株式会社 日立製作所  
【代理人】  
【識別番号】 100075096  
【弁理士】  
【氏名又は名称】 作田 康夫  
【手数料の表示】  
【予納台帳番号】 013088  
【納付金額】 21,000円  
【提出物件の目録】  
【物件名】 特許請求の範囲 1  
【物件名】 明細書 1  
【物件名】 図面 1  
【物件名】 要約書 1

**【書類名】特許請求の範囲****【請求項 1】**

ネットワークに接続される記憶装置であって、  
前記ネットワークと接続され、かつファイルアクセスを受け付けるインターフェースと  
複数のディスクドライブと、  
前記ファイルアクセスをロックアクセスへ変換し、前記ロックアクセスに基づいて  
前記複数のディスクドライブを制御する制御部とを有し、  
前記制御部が、  
前記インターフェース、前記複数のディスクドライブ及び前記制御部を論理的に分割し  
複数の仮想的な記憶装置として独立に動作させることを特徴とする記憶装置。

**【請求項 2】**

前記制御部は更にキャッシュメモリを有し、前記キャッシュメモリが前記複数の仮想的  
な記憶装置の各々に論理的に分割されて割り当てられていることを特徴とする請求項 1 記  
載の記憶装置。

**【請求項 3】**

前記制御部は更に前記ファイルアクセスを前記ロックアクセスへ変換する第一のプロ  
セッサと前記ロックアクセスに基づいて前記複数のディスクドライブを制御する第二の  
プロセッサとを有し、

前記第一のプロセッサ及び前記第二のプロセッサが各々論理的に分割され、前記複数の  
仮想的な記憶装置の各々に割り当てられていることを特徴とする請求項 2 記載の記憶装置  
。

**【請求項 4】**

前記第一のプロセッサが、前記インターフェース及び該第一のプロセッサの論理分割を  
行う第一のハイパバイザを実行し、

前記第二のプロセッサが、前記キャッシュメモリ、複数のディスク装置及び該第二のプロ  
セッサの論理分割を行う第二のハイパバイザを実行することを特徴とする請求項 3 記載  
の記憶装置。

**【請求項 5】**

前記制御部はさらに、前記第一のプロセッサが使用するメモリ及び前記第一のプロセッ  
サと前記第二のプロセッサとを接続する通信網を有し、

前記メモリは前記第一のハイパバイザによって論理的に分割され、前記通信網は前記第  
二のハイパバイザによって論理的に分割されることを特徴とする請求項 4 記載の記憶装置  
。

**【請求項 6】**

前記第一のプロセッサ及び前記第二のプロセッサが、前記インターフェース、前記第一  
のプロセッサ、前記キャッシュメモリ、前記第二のプロセッサ及び前記複数のディスクド  
ライブの論理分割を行うハイパバイザを実行することを特徴とする請求項 3 記載の記憶裝  
置。

**【請求項 7】**

前記制御部が、前記インターフェース、該制御部及び前記複数のディスクドライブの論  
理分割を行うハイパバイザを実行することを特徴とする請求項 1 記載の記憶装置。

**【請求項 8】**

更に管理端末と接続され、

前記制御部は、前記管理端末から入力された情報に基づいて前記論理分割を行うことを  
特徴とする請求項 3 記載の記憶装置。

**【請求項 9】**

前記管理端末に入力される情報が、該記憶装置を使用する上位装置がデータ転送速度を  
重視するという情報であれば、前記複数の仮想的な記憶装置のうち、前記上位装置が使用  
する仮想的な記憶装置に対する前記キャッシュメモリの割り当てを増加させることを特徴

とする請求項8記載の記憶装置。

**【請求項10】**

前記管理端末に入力される情報が、該記憶装置を使用する上位装置が広範囲のランダムなアクセスをするという情報であれば、前記複数の仮想的な記憶装置のうち前記上位装置が使用する仮想的な記憶装置に対する前記キャッシュメモリの割り当てを減少させることを特徴とする請求項8記載の記憶装置。

**【請求項11】**

更に管理端末と接続され、

前記制御部は、前記管理端末から入力された情報に基づいて前記論理分割を行うことを特徴とする請求項5記載の記憶装置。

**【請求項12】**

前記管理端末に入力される情報が、該記憶装置を使用する上位装置が逐次的な連続したアクセスをするという情報であれば、前記複数の仮想的な記憶装置のうち前記上位装置が使用する仮想的な記憶装置に対する前記キャッシュメモリ及び前記メモリの割り当てを増加させることを特徴とする請求項11記載の記憶装置。

**【請求項13】**

前記管理端末に入力される情報が、該記憶装置を使用する上位装置が少数の大容量のファイルをアクセスをするという情報であれば、前記複数の仮想的な記憶装置のうち前記上位装置が使用する仮想的な記憶装置に対する前記第一のプロセッサの割り当て量を減少させ、前記第二のプロセッサの割当量を増加させることを特徴とする請求項8記載の記憶装置。

**【請求項14】**

前記管理端末に入力される情報が、該記憶装置を使用する上位装置が多数の小容量のファイルをアクセスをするという情報であれば、前記複数の仮想的な記憶装置のうち前記上位装置が使用する仮想的な記憶装置に対する前記第一のプロセッサの割り当て量を増加させ、前記第二のプロセッサの割当量を減少させることを特徴とする請求項8記載の記憶装置。

**【請求項15】**

前記管理端末に入力される情報が、該記憶装置を使用する上位装置が大容量のファイルを逐次的にアクセスするという情報であれば、前記複数の仮想的な記憶装置のうち前記上位装置が使用する仮想的な記憶装置に対する前記通信網の論理的な割当量を減少させることを特徴とする請求項11記載の記憶装置。

**【請求項16】**

ネットワークと接続され、かつファイルアクセスを受け付けるインターフェースと、複数のディスクドライブと、前記ファイルアクセスをロックアクセスへ変換し、前記ロックアクセスに基づいて前記複数のディスクドライブを制御する制御部とを有する記憶装置と、

前記記憶装置と接続される管理端末を有する記憶装置システムであって、

前記記憶装置は、前記管理端末に入力される情報に基づいて、前記インターフェース、前記複数のディスクドライブ及び前記制御部を論理的に分割し、複数の仮想的な記憶装置として独立に動作することを特徴とする記憶装置システム。

**【請求項17】**

前記管理端末に入力される情報は、前記記憶装置を使用する計算機のアクセス特性についての情報であり、前記記憶装置は、前記管理端末に入力される前記アクセス特性についての情報に基づいて前記記憶装置が有する資源の論理分割量を計算し、その結果を用いて前記論理的な分割を行うことを特徴とする請求項16記載の記憶装置システム。

【書類名】明細書

【発明の名称】論理分割可能な記憶装置及び記憶装置システム

【技術分野】

【0001】

本発明は、ネットワークに接続して用いる記憶装置、特にNASに関する。

【背景技術】

【0002】

情報処理システムの能力を向上するための方法として、単純に情報処理システムが有する計算機の台数を増やす方法がある。しかし、計算機を多数設置すると個々の計算機の管理に手間がかかり、またこれらの複数の計算機の設置面積や電力等の合計も非常に大きくなるという問題があった。これを解決するために、処理能力の大きい計算機を設置し、その計算機の資源を論理的に複数に分割し、その分割された部分の各々を仮想的な計算機として各々独立に使用する技術が考案されている。これを計算機の論理分割(Logical Partitioning: LPAR)と呼ぶ。例えば特許文献1にこのような論理分割技術の例が開示されている。

【0003】

論理分割によって一つの計算機を多数の計算機に仮想的に見せることにより、仮想的な計算機の各々に種々のオペレーティング・システムを自由に搭載し、運用・停止や障害処理も仮想的な計算機毎に独立して行えるなど、柔軟な運用が可能となる。また物理的な装置台数が少ないので、装置管理、設置面積、電力等で有利である。但し、従来の計算機におけるLPARでは、計算機内のプロセッサ、メモリなどの資源が論理分割されてそれぞれの仮想的な計算機に割り付けられていたが、計算機に接続される記憶装置については、その記憶装置が有する記憶領域が分割されて仮想的な計算機に各々割り当てられるだけで、それ以上の考慮は特にされていなかった。

【0004】

一方、記憶装置には、1台の計算機(以下「ホスト」と称することもある)に直結して用いられる形態のほかに、ネットワークを介して複数の計算機に共有される形態がある。この形態において特に、ファイルシステムのインターフェースを有する、すなわち計算機からファイルアクセスが可能な記憶装置をネットワークアタッチドストレージ(以下「NAS」と称する)。

【0005】

NASとホストとのデータのやり取りは、ホスト上で動作するオペレーティング・システムが認識する名前や構造を持ったファイルという単位で行われる。このため、NASは、データを記憶するためのディスクドライブやその制御部に加え、ホストとのファイル入出力をディスクドライブとのデータ入出力に変換するために使用するプロセッサやメモリを有する。

【0006】

【特許文献1】特開2003-157177号公報

【発明の開示】

【発明が解決しようとする課題】

【0007】

NASは、もともと複数のホストにそれぞれ記憶装置を別個に設置するよりネットワークで共有する記憶装置を設けると有利であるという思想に基づいている。このため、複数のホスト間で記憶領域等を共有するための制御をNAS自身が行わなければならなかつた。またあるホストが大量にデータの読み書きを行うと、NASの処理能力の大半が消費され、他のホストに対するデータの入出力の能力が低下する。更に、あるホストの誤操作等でNASのデータ破壊や障害が発生すると、他のホストが使用するデータにも影響を及ぼす場合もあった。

【0008】

本発明が解決しようとする課題は、NASが複数のホストから共有されるときに、共有

にかかる制御を削減するとともに、相互の計算機の干渉を排除し、入出力の能力の保証・データ破壊や障害の局所化を行うことである。また本発明が解決しようとする別の課題は、NAS内のプロセッサ、メモリ等の資源の利用率を高めることである。

【課題を解決するための手段】

【0009】

上記の課題を解決するために、本発明では、NASにおいて論理分割を行う。より具体的には、ネットワークに接続され、ファイルアクセスを受け付ける記憶装置において、記憶装置が有する各資源、例えばディスクドライブ、ネットワークとのインターフェース、ファイルアクセスを制御するプロセッサ等を、記憶装置が有する制御部が論理的に分割して、個々の論理区画（仮想的な記憶装置）を独立して動作させる構成とする。

【0010】

尚、制御部は複数のプロセッサを有し、これらのプロセッサが論理分割を分担して行ったり、全体として行ったりしても良い。

更に、管理端末を有し、論理分割に必要な情報をこの管理端末から入力する構成としても良い。この場合、管理端末には記憶装置を使用する計算機のアクセス特性のみが入力され、管理端末がそのアクセス特性から論理分割に必要な情報を算出して記憶装置に伝達する構成としても良い。

更に、記憶装置を使用する計算機が管理端末を兼ねる構成でも良い。

【発明の効果】

【0011】

本発明による記憶装置では、複数のホストから共有されるときに、共有にかかる制御を削減するとともに、ホスト相互の干渉を排除し、データの入出力の能力の保証・データ破壊や障害の局所化を行うことができる。

又、ホスト毎またはホストのグループ毎に独立したユーザ認証を行うことができる。また本発明により、記憶装置内のプロセッサ、メモリ、記憶媒体等の資源の利用率を高めることができることが可能である。

【0012】

更に、1台のNASで複数の仮想的なNASを提供することができ、ホストのオペレーティング・システムの自由度が増し、運用・停止や障害処理も独立して行うことが可能であり、装置管理、設置面積、電力等で有利となる。

【発明を実施するための最良の形態】

【0013】

以下、本発明の実施の形態を図面を用いて説明する。尚、本発明が下記の実施形態の記載に限定されることが無いのは言うまでもない。

図1は、本発明が適用されたNASの実施形態一例を示す図である。NASは、ホスト（以下「上位装置」とも称する）と接続される二つのホストインターフェース部100、ホストからのファイル単位の入出力要求をロック単位のデータの入出力要求に変換する三つのファイル制御プロセッサ101、ファイル単位の入出力要求をロック単位のデータの入出力要求に変換するために必要な情報（以下「変換情報」とも称する）を格納する二つのファイル制御メモリ102、記憶媒体である四つのディスクドライブ群107、ディスクドライブ群107とのデータの入出力の制御を行う三つのディスク制御プロセッサ104、ディスクドライブ群107に入出力されるデータを一時的に蓄える二つのキャッシュメモリ106、ディスク制御プロセッサ104とディスクドライブ群107とを接続する二つのディスクIF部106、ディスク制御プロセッサ104とファイル制御プロセッサ101とを接続する二つのプロセッサ間通信接続部103を有する。

【0014】

ここで、ディスクドライブ群107には、複数のディスクドライブが含まれ、各グループ毎にRAID構成がとられている場合もある。また、「ロック」とは、ディスク制御プロセッサ104がディスクドライブへデータを格納する所定の単位であり、一般的に512Bが採用される。又、「変換情報」とは、ファイルシステムで使用されるファイル名及びファイル

の先頭からの位置とブロックとの対応関係を示す情報であり、一般的にはI-Node等のリンク構造や、アドレス変換テーブルのようなデータ構造で表現される。

【0015】

尚、本明細書においては、同一の装置には同一の番号を付し、同一の装置を区別するときは英字のa、b等を番号に付与する。又、上述した各装置の個数も例示であり、本発明を限定するものではない。

【0016】

図1において、上位装置からNASに送信されたファイル単位の入出力要求は以下のようにNASで処理される。

NASにはまず初めに、ホストからファイル名を指定したファイル使用開始（オープン）処理の要求が送信される。次に実際のデータの入出力要求が送信され、最後にファイル使用終了（クローズ）処理の要求が送信される。

【0017】

これらの要求は、いずれかのホストインターフェース部100で受信され、いずれかのファイル制御プロセッサ101に渡される。ファイル制御プロセッサ101は、ファイル制御メモリ102に格納された変換情報を参照して、ホストに要求されたファイル名を確認し、使用開始されたファイル名を記録し、該ファイルに対するデータ入出力要求を、データが格納されているディスクドライブ群107へのデータ入出力要求に変換する。

【0018】

変換されたデータ入出力要求は、いずれかのプロセッサ間通信接続部103を介して、いずれかのディスク制御プロセッサ104に送出される。またファイル制御メモリ102に必要な変換情報が格納されていない場合は、ファイル制御プロセッサ101は、プロセッサ間通信接続部103を介して、いずれかのディスク制御プロセッサ104にディスクドライブ群107の予め定められた記憶領域に格納されている変換情報を要求する。

【0019】

ディスク制御プロセッサ104は、いずれかのファイル制御プロセッサ101よりプロセッサ間通信制御部103を介して受け取ったデータ入出力要求（変換情報の要求も含む）に対し、当該データがいずれかのキャッシュメモリ105に格納されていないかどうか調べる。当該データがいずれかのキャッシュメモリ105に格納されていた時には、ディスク制御プロセッサ104は、要求されたデータの書き込み又は読み出しをキャッシュメモリ105に対して行う。

【0020】

その後、ディスク制御プロセッサ104は、書き込みの時には完了したという結果を、読み出しの時には読み出しの完了と読み出したデータとを合わせて、プロセッサ間通信部103を介して、入出力要求を送信したファイル制御プロセッサ101に返送する。ファイル制御プロセッサ101は、返送された結果及びデータを処理し、入出力要求を受信したホストインターフェース部100を介して、NASに入出力要求を送信した上位装置に結果（データあるいは処理完了の報告等）を送信する。

【0021】

一方、ファイル制御プロセッサ101から要求されたデータが全てのキャッシュメモリ105に格納されていなかった場合、ディスク制御プロセッサ104は、要求されたデータがディスクドライブ群107のどの部位に格納されているかを特定し、いずれかのディスクIF部106を介して、ディスクドライブ群107よりデータを読み出して、いずれかのキャッシュメモリ105に格納する。

【0022】

その後、ディスク制御プロセッサ104は、要求されたデータの読み出し又は書き込みをデータが格納されたキャッシュメモリ105に対して行う。それ以降の処理は、上述と同様である。

【0023】

尚、キャッシュメモリ105に格納されているデータは、一定時間経過又はキャッシュメ

モリ105の空き領域が不足した時などにディスクドライブ群107に書き戻される。

【0024】

本実施形態において、上述した処理は、例えば図1に示したような論理分割境界によって分割された単位（論理区画）によって、それぞれ独立して行われる。それぞれの論理区画に割り当てられる処理のための物理的な資源であるホストインターフェース部100、ファイル制御プロセッサ101、ファイル制御メモリ102、プロセッサ間通信接続部103、ディスク制御プロセッサ104、キャッシングメモリ105、ディスクIF部106及びディスクドライブ群107は、各論理区画に一度割り当てられるとその論理区画の処理に専ら用いられる。具体的に言えば、図1において違う論理区画に割り当てられたファイル制御プロセッサ101aとディスク制御プロセッサ104cとは、上述したデータの入出力要求の取り扱いを行わない。

【0025】

また図1で論理分割境界をまたがっている表示されている資源（例えばファイル制御メモリ102b）は、その容量等が予め割り当てられた比率で論理的に分割されて各論理区画毎に用いられる。このようにすることにより、各論理区画がそれぞれ独立した仮想的なNASとして動作する。

【0026】

各論理区画への物理資源の論理的な分割・割り当て処理は、実際にはファイル制御プロセッサ101やディスク制御プロセッサ104で実行される。論理分割の制御の方法としては、以下の二つの方法が考えられる。

【0027】

一つ目は、ファイル制御プロセッサ101とディスク制御プロセッサ104が幾つかの物理資源の論理分割の制御を分担して行い、全体としてはお互いが連携して論理分割を制御する方法である。

例えば、ホストインターフェース部100、ファイル制御プロセッサ101及びファイル制御メモリ102の割り当て処理を、ファイル制御プロセッサ101が行う。この処理を以下「ファイル制御ハイパバイザ」と呼ぶ。

【0028】

またプロセッサ間通信接続部103、ディスク制御プロセッサ104、キャッシングメモリ105、ディスクIF部106及びディスクドライブ群107の割り当て処理を、ディスク制御プロセッサ104が行う。この処理を以下「ディスク制御ハイパバイザ」と呼ぶ。ファイル制御プロセッサ101で実行されるファイル制御ハイパバイザと、ディスク制御プロセッサ104で実行されるディスク制御ハイパバイザは、互いに連携して、それぞれの割り当て処理を行う。連携の具体的な内容については、後述する。尚、ファイル制御ハイパバイザを実行するファイル制御プロセッサ101はいずれか一つ、例えばファイル制御プロセッサ101aのみであってもよく、又は複数、例えばファイル制御プロセッサ101aおよびbであっても良い。これはディスク制御ハイパバイザでも同様である。

【0029】

二つ目は、二つの制御プロセッサが共同して全体の物理資源の論理分割を制御する方法である。具体的には、NASの全ての資源の論理区画の割り当て処理（以下「統合ハイパバイザ」）を、ファイル制御プロセッサ101a～101c、ディスク制御プロセッサ104a～104cのすべてで行う。

【0030】

具体的には、各プロセッサ上で動作するハイパバイザは、例えば次のようにして論理分割を実現する。

まず、各プロセッサ上で動作する基本I/O処理ソフトウェア（BIOS）に対し、当該プロセッサが割り当てられた論理区画内のI/O処理資源以外の資源が見えないようにする。例えば、図1でファイル制御プロセッサ101aは物理的にはホストインターフェース部100bと接続されているが、図示されている点線で論理区画が分かれる設定がされた場合は、ホストインターフェース部100bを見えないようにする。

**【0031】**

より具体的には、BIOS内で当該プロセッサに接続されている資源や利用可能な資源を調べるための特権命令が実行された場合、当該特権命令の実行でソフトウェア的な割り込みを発生させ、ハイパバイザに実行を移す。ハイパバイザ内で当該プロセッサが属している論理区画に対して割り当てられている資源を調べ、その論理区画に割り当てられている資源のみが接続されているように当該特権命令の結果をセットし、割り込みを発生したBIOSへ復帰する。

**【0032】**

このようにすることにより、各プロセッサはそのプロセッサの属する論理区画の資源のみを扱うことになり、論理区画間の分離が実現される。

資源の中にはメモリや複数の通信チャネルを備えるホストインターフェース部100やプロセッサ間通信部103があるが、このような場合は、それぞれの論理区画のプロセッサに見せる資源の量（メモリなら開始物理アドレス、終了物理アドレスでメモリの容量、通信チャネルならチャネルの物理番号の組で示されるチャネル数）を制御すればよい。

**【0033】**

またプロセッサ自身については、それぞれのプロセッサが一つの論理区画に完全に割り当てられている場合には、当該プロセッサはその論理区画の処理で占有させれば良い。

一方、ある一つのプロセッサを二つ以上の論理区画に割り当てて各々の共有割合を定めて共有させることも考えられる。このような場合は、タイマー割り込みをハードウェアで各プロセッサに実装し、そのタイマー割り込みでハイパバイザが一定時間ごとに起動されるようにする構成が考えられる。

**【0034】**

上述のタイマー割り込みで起動されたハイパバイザは、当該プロセッサでそれぞれの論理区画の処理をどのくらい行ったかを計測し、予め定めた共有割合に従って次に処理を行うべき論理区画を決定して、その論理区画の処理へプロセッサの実行を移す。このようにすれば、一つのプロセッサを時間的に予め定めた割合に分割して、二つ以上の論理区画に割り当てることができる。

**【0035】**

なお、ハイパバイザの実現方法としては、上記の例のほかにも、例えば、各プロセッサに接続され、資源管理を行う専用のハードウェアや、小規模のマイクロプログラムで制御されるような専用のコプロセッサ等を搭載して、論理分割の制御を実現してもよい。

**【0036】**

又、論理分割に関する情報、例えば論理区画1で使用されるプロセッサ、メモリ、通信接続部等を指定する情報等は、ファイル制御メモリ102、キャッシュメモリ105、ディスク群107のディスクドライブ又はその他の記憶媒体のいずれか一つ又は複数に格納されており、各ハイパバイザは、その情報を読み出すことによって、論理分割の指定をBIOS等に対して行う。尚、この情報は、後述する管理端末等を介して設定される。

**【0037】**

上記のようにして実現されたハイパバイザについて、先に述べたように、ファイル制御プロセッサでファイル制御ハイパバイザ、ディスク制御プロセッサでディスク制御ハイパバイザを動作させる場合は、ファイル制御ハイパバイザは、ホストインターフェース部100、ファイル制御プロセッサ101及びファイル制御メモリ102の割り当て処理を、ディスク制御ハイパバイザは、プロセッサ間通信接続部103、ディスク制御プロセッサ104、キャッシュメモリ105、ディスクIF部106及びディスクドライブ群107の割り当て処理を行い、二種類のハイパバイザを連携させるようにする。

**【0038】**

具体的には、後述するような管理端末等によって論理分割の指定を行う際に、ファイル制御ハイパバイザに対する指定とディスク制御ハイパバイザに対する指定を関連付けて行われるようにする。あるいは、論理分割に与える指定の仕方によっては、後述するように、ファイル制御ハイパバイザとディスク制御ハイパバイザが、当該論理区画が指定された

論理分割に関する要求に合致するように、互いに自動的に調整し合うようとする。

【0039】

統合ハイパバイザの場合には、例えば、統合ハイパバイザを起動する各プロセッサが、すべての資源の各論理区画への割り当て情報を共有し、当該ハイパバイザが、起動したプロセッサが用いる資源をその割り当て情報を参照して決定し、割り当て処理を実行する。

【0040】

次に、NASの各資源の論理区画への割り当ての具体例を説明する。以下、図1で示したNASを二つの論理区画（論理区画1及び論理区画2）に論理分割した場合を例として説明する。しかし、論理区画の数は幾つでも構わない。又、以下ではファイル制御ハイパバイザとディスク制御ハイパバイザが連携して論理分割を行うとして説明するが、統合制御ハイパバイザでも構わない。更に、ハイパバイザを主語とした場合は、その処理は実際は各ハイパバイザの処理を行うプロセッサによって実行されているものとする。

【0041】

図2は、ファイル制御メモリ102及びキャッシュメモリ105の論理区画への割り当ての例を示す図である。例えば論理区画1を使用する上位装置の要求がデータ読み出し速度重視であるときは、ディスク制御ハイパバイザは、論理区画1へのキャッシュメモリ105の割り当て量を増やし、要求されたデータができるだけキャッシュメモリ105に格納されるようとする。このときはファイル制御メモリ102の論理区画1への割り当て量は少なくてもよい。

【0042】

これに対応して、ファイル制御ハイパバイザは、論理区画1へのファイル制御メモリ102の割り当て量を下げ、論理区画2により多く記憶容量を割り当てる。これにより、NAS全体でのファイル制御メモリ102の利用率を向上させることが可能である。

【0043】

一方、論理区画1を使用する上位装置の要求が応答速度重視であれば、変換情報がなるべくファイル制御メモリ102に格納されるように、ファイル制御ハイパバイザは、論理区画1により多くのファイル制御メモリ102の記憶容量を割り当てる割り当て処理を行う。このときは、論理区画1へのキャッシュメモリ105の割り当て量は少なくてもよい。この場合、ディスク制御ハイパバイザは論理区画2により多くのキャッシュメモリ105の記憶容量を割り当てることができ、NAS全体として、キャッシュメモリ105の利用率を向上することが可能である。

【0044】

又、論理区画1を利用する上位装置の入出力要求が、NASが有するディスクドライブ群107の広範囲に散らばるデータへのランダムなアクセスが主体の場合、論理区画1にファイル制御メモリ102及びキャッシュメモリ105の多くの記憶容量を割り当ても広範囲に散らばるアクセスのすべての情報を格納することは難しいので、その割り当ての効果が薄い。従って、このような場合には、ファイル制御ハイパバイザとディスク制御ハイパバイザがファイル制御メモリ102及びキャッシュメモリ105の論理区画1への割り当て量を少なくし、他の論理区画である論理区画2に記憶容量を多く割り当て、キャッシュメモリ105等の利用率を向上させる。

【0045】

逆に論理区画1を利用する上位装置の入出力要求が、NASが有するディスクドライブ群107の連続範囲に格納されたデータへのシーケンシャルなアクセスが主体の場合、NAS自体でアクセスに必要な情報及び後に読み出されるデータを前もって特定することができる。そのため、それらの情報やデータを充分ファイル制御メモリ102及びキャッシュメモリ105に格納することができるように、ファイル制御ハイパバイザ及びディスク制御ハイパバイザが、ファイル制御メモリ102及びキャッシュメモリ105の論理区画1への割り当て量を増やすように割り当てる処理を行うことが考えられる。

【0046】

図3は、ファイル制御プロセッサ101及びディスク制御プロセッサ104の論理区画への割

り当ての例を示す図である。

論理区画 1 を使用する上位装置からの入出力要求が少数の大容量ファイルアクセスの場合、ファイル制御プロセッサ101で実行されるファイル入出力をデータ入出力に変換する処理の量はあまり多くない。従って、ファイル制御プロセッサ101の論理区画 1 への割り当て量は少なくてよい。

#### 【0047】

この場合、ファイル制御ハイパバイザが論理区画 1 へのファイル制御プロセッサ101の論理区画 1 への割り当て量（具体的にはプロセッサの占有率）を下げ、他の論理区画である論理区画 2 に論理区画 1 に比較して相対的に多くプロセッサ資源を割り当てるにより、NAS が有するファイル制御プロセッサ101の利用率の向上が可能になる。

#### 【0048】

又この場合は、ファイルのデータ量が大きいため、論理区画 1 に割り当てられたディスク制御プロセッサ104で実行されるデータ入出力の処理の量は多くなる。したがって、ディスク制御ハイパバイザは、論理区画 1 へのディスク制御プロセッサ104の割り当て量を増やすように割り当て処理を行う。

#### 【0049】

また、論理区画 1 を使用する上位装置の入出力要求が多数の小容量ファイルアクセスである場合、ファイル制御プロセッサ101で実行されるファイル入出力をデータ入出力に変換する処理の量は多くなる。そこで、ファイル制御ハイパバイザは、論理区画 1 へのファイル制御プロセッサ101の割り当て量を増やすように割り当て処理を行う。

#### 【0050】

このとき、ファイルのデータ量自体は小さいので、論理区画 1 に割り当てられたディスク制御プロセッサ104で実行されるデータ入出力の処理の量はあまり多くない。そこで、ディスク制御ハイパバイザは、論理区画 1 へのディスク制御プロセッサ104の割り当て量を減らして、論理区画 2 へのディスク制御プロセッサ104の割り当て量を増やす。これにより、NAS におけるディスク制御プロセッサ104の利用率の向上が可能となる。

#### 【0051】

さらに、論理区画 1 を使用する上位装置が高性能のNAS を必要としない場合は、ファイル制御ハイパバイザ及びディスク制御ハイパバイザは、ファイル制御プロセッサ101及びディスク制御プロセッサ104の論理区画 1 への割り当て量を減らすように割り当て処理を行う。逆に、論理区画 1 を使用する上位装置が高性能のNAS を必要とする場合は、ファイル制御ハイパバイザ及びディスク制御ハイパバイザは、ファイル制御プロセッサ101及びディスク制御プロセッサ104の論理区画 1 への割り当て量を増やすように割り当て処理を行う。

#### 【0052】

図4は、プロセッサ間通信接続部103の論理区画への割り当ての例を示す図である。論理区画 1 を使用する上位装置からの入出力要求が大容量のシーケンシャルアクセスである場合、ディスク制御ハイパバイザは、論理区画 1 へのプロセッサ間通信接続部103の割り当て量（具体的には通信帯域）を増やし、ファイル制御プロセッサ101とディスク制御プロセッサ104との間のデータ通信能力（言い換えると、ホストインターフェース部100からキャッシュメモリ105までの間のデータ通信能力）を確保するように割り当て処理を行う。

#### 【0053】

また、論理区画 1 を使用する上位装置の入出力要求が小容量のシーケンシャルアクセスであれば、論理区画 1 へのプロセッサ間通信接続部103の割り当て量は大きくなくてもよい。さらに上位装置からの要求がランダムアクセスであれば、上位装置からみた論理区画 1 で構成される仮想的なNAS の性能はプロセッサ間通信接続部103の論理区画 1 への割り当て量にあまり影響されない。従ってこれらの場合は、ディスク制御ハイパバイザは、論理区画 1 へのプロセッサ間通信接続部103の割り当て量を減らして他の論理区画（ここでは論理区画 2 ）へ割り当てを増やし、NAS におけるプロセッサ間通信接続部103の利

用率を高めるように割り当て処理を行う。

【0054】

図5は、ディスクドライブ群107の論理区画への割り当ての例を示す図である。論理区画1を使用する上位装置が記憶容量優先であれば、ディスク制御ハイパバイザは、記憶容量効率の高いRAID5構成（図7ではデータが格納されているディスクドライブ3個に対しパリティが格納されているディスクドライブ1個の割合で容量効率は75%となる）になっているディスクドライブ群701を論理区画1に割り当てるように処理を行う。このとき、ディスクドライブの回転速度も例えば毎分7,500回転など、あまり速いものでなくともよい。

【0055】

一方、論理区画1を使用する上位装置がアクセス性能を重視する場合であれば、ディスク制御ハイパバイザは、アクセス性能を高められるRAID1構成（図7において、同一のデータが複製されて2つのディスクドライブに格納されるので、記憶容量効率は50%だが、同一のデータに対して2個のディスクドライブのいずれも使用可能なので、トータルのアクセス性能は一つのディスクドライブの2倍になる）になっているディスクドライブ群701を論理区画1に割り当てるように処理を行う。尚、この場合、ディスク制御ハイパバイザは、ディスクドライブ群107に含まれるディスクドライブの回転速度も考慮して、同じRAID1構成のディスクドライブ群107のうち、高回転速度、例えば毎分15,000回転のディスクドライブを有するディスクドライブ群107を論理区画1に割り当てる処理を行っても良い。

【0056】

尚、ホストインターフェース部100の論理区画への割り当ては、各論理区画を使用する上位装置から論理区画に要求される性能に応じて、ファイル制御ハイパバイザが割り当てる。具体的には、上位装置の要求性能が高い場合には、ファイル制御ハイパバイザは、上位装置が使用する論理区画に大きな割り当て量、即ち上位装置との間の高い通信能力（通信帯域等）を割り当てる。一方、上位装置の要求性能が低いあるいは特に要求が無い場合には、ファイル制御ハイパバイザは、上位装置が使用する論理区画に小さな割り当て量、即ち上位装置との間の低い通信能力（通信帯域等）を割り当て、NAS全体の効率を重視することが考えられる。

【0057】

更に本実施形態のように1つのNASを論理的に分割して用いることにより、NASにおける上位装置のユーザ認証を、論理区画単位毎に独立して行うことが可能である。図6はその概念の例を示す図である。

本図では、論理区画1を使用する上位装置Aを識別子（以下「ID」）= “abc”であるユーザAとID= “def”であるユーザBが使用し、論理区画2を使用する上位装置BをID= “ghi”であるユーザCとID= “abc”であるユーザDが使用する。このとき上位装置AのユーザAと上位装置BのユーザDが同じID= “abc”であるので、このユーザAとユーザDをNASで区別して扱うためには、従来は上位装置または上位装置のグループにIDを付与してそのIDとユーザのIDを組み合わせてユーザを区別するなどの特別な処理が必要であった。

【0058】

しかし本実施形態では、ホストインターフェース部100及びファイル制御プロセッサ101が論理的に分割され、各論理区画が別々の仮想的なNASのホストインターフェース部100及びファイル制御プロセッサ101として動作するので、ユーザ認証も各論理区画毎に独立して行われる。即ち、図8において同一のID= “abc”を持つユーザAとユーザDは、別々の論理区画でユーザ認証されるので、自然に区別して扱われ、ユーザAとユーザDを区別するために特別な処理をなんら必要としない。つまり、論理区画さえ異なれば、特別な処理をすることなく、複数のユーザに同一のIDを与えることができる。

【0059】

さらには、各論理区画に割り当てられているホストインターフェース部100及びファイル制御プロセッサ101の資源は他の論理区画で使用されることはないので、ある論理区画

のユーザが大量のデータアクセスを行っても、ほかの論理区画のユーザは影響を受けることがない。

#### 【0060】

次に、第二の実施形態について説明する。本実施形態のNASは、上述したNASの、ファイル制御プロセッサ101及びディスク制御プロセッサ104並びにファイル制御メモリ102及びキャッシュメモリ105を統合した、それぞれ一種類のプロセッサ、メモリを有する。

図7は、第二の実施形態の構成の例を示す図である。図9において、統合制御プロセッサ901は、ファイル制御プロセッサ101及びディスク制御プロセッサ104を統合したプロセッサであり、統合メモリ902は、ファイル制御メモリ102及びキャッシュメモリ105を統合したメモリである。

#### 【0061】

上述した実施形態(図1)に比べ、本実施形態では、プロセッサ間通信接続部103が不要となり、装置構成が簡単になる。図7において、上位装置からのファイル単位の入出力要求をデータの入出力要求に変換する処理と、ディスクIF部106とディスクドライブ群107との間のデータの入出力の制御は、両方とも統合プロセッサ901で行われる。また、統合メモリ902には、変換情報及びディスクドライブ群107のデータが格納される。図7における他の部分の構成・動作は図1と同じである。

#### 【0062】

図7においても、図1と同様に、例えば図7中に示したような論理分割境界によって分割された論理区画で、それぞれ独立して処理が行われる。それぞれの論理区画に割り当てられる処理のための物理的な資源であるホストインターフェース部100、ホストインターフェース部100、統合制御プロセッサ901、統合メモリ902、ディスクIF部106及びディスクドライブ群107は、各論理区画に一度割り当てられるとその論理区画の処理に専ら用いられる。このようにすることにより、各論理区画がそれぞれ独立した仮想的なNASとして動作する。

#### 【0063】

本実施形態における各論理区画への物理資源の論理的な分割・割り当て処理は、実際には統合制御プロセッサ901が実行する。統合制御プロセッサ901は、上述した統合制御ハイパバイザの制御を行う。

#### 【0064】

図8は、NASの論理分割の設定を入力するための管理端末の入力画面の例を示す図である。このような入力画面は、第一の実施形態及び第二の実施形態の双方で使用される。管理者等が図8のような論理分割の設定入力をやってその内容がNASに通知されることによって、NASで動作するハイパバイザがNASの各資源を論理分割する。より詳細に言えば、設定入力の内容が、NASが有するある記憶領域に格納され、ハイパバイザは、この格納された情報にしたがって論理分割を行う。

#### 【0065】

このような管理端末は、具体的にはネットワークを通じてNASに接続する上位装置が有していてもよい。また、NASに専用線で接続されるコンソール装置で実現されていてもよい。このようなコンソール装置はキーボード等の入力装置とディスプレイ等の表示装置で実現することができる。

#### 【0066】

管理者等が入力した情報は、上位装置やコンソール装置から専用のプロトコル又は汎用のプロトコルを用いてNASに転送される。NASはその情報を受取るためのインターフェース(例えばホストインターフェース部100又は専用のインターフェース)を有する。

#### 【0067】

以下、図8で示される設定入力画面について詳細に説明する。該画面には論理分割数を入力するフィールドがあり、管理者等は、初めにNASの論理分割数をいくつにするかをこのフィールドに入力する(図8の例では3)。この際、管理者等が論理分割数を入力すると、その数に応じた論理区画が物理資源毎に画面に表示され、各論理区画への資源割り当

ての初期値が表示される。

【0068】

その後、管理者等は、各プロセッサ、メモリ等の資源の分配を画面を見ながら入力していく。このとき、例えばファイル制御プロセッサ101とディスク制御プロセッサ104の割り当てを指定する部分では、図示するように、それぞれを各論理区画にどのように割り当てるかを関連して設定できるような表示（図ではプロセッサ同士を並べて表示し、関連性がわかりやすいようにしている）にすれば、先に図5で説明したような割り当て制御の設定を入力することが容易になる。

【0069】

また例えば図示したように、ファイル制御プロセッサ101やディスク制御プロセッサ104の論理区画に対するそれぞれの割り当て量を個別に設定するつまみ（ポインティングデバイスで選択できる部分）のほかに、連動して設定することができるつまみも用意しておく。同様にファイル制御メモリ102とキャッシュメモリ105についても、各論理区画への割り当て量を関連づけて表示し、個別設定・連動設定のつまみを用意しておく。

【0070】

図8の例では、管理者等は、プロセッサ間通信接続部103について、ファイル制御プロセッサ101とディスク制御プロセッサ104（ホストインターフェース部100とキャッシュメモリ105）の間の全体のデータ転送能力を、それぞれ各論理区画にどの割合で割り当てるかを入力する。

【0071】

又、管理者等は、ホストインターフェース部100について、資源の割合ではなく物理的に上位装置を接続するためのネットワークの接続口（ポート）を特定することで割り当てる情報を入力する。ただし、単に資源の割合で入力する入力方法でも良い。更に管理者等は、ディスクドライブ群107について、各論理区画に割り当てる物理的なディスクドライブの容量・RAID構成・能力（回転数）を設定することで資源の割り当てを行う。

【0072】

以上の割り当て設定の入力は例であり、これ以外にも、個別に数値入力することやある程度の自動設定をしてもよい。例えば、上位装置がある論理区画に要求するアクセスの特性（ランダムかシーケンシャルか、1転送あたりの平均的なデータ長、最小データ転送速度、最大レスポンス時間等）を管理者等が管理端末を介して入力することにより、予めいくつか作成してある設定値のセット、具体的には図2～5で示された特徴を有する設定値のセットから入力された特性に合うものをそれぞれハイパバイザで選んで設定するようにしてもよい。

【0073】

これにより、例えば、管理者がシーケンシャルを指定した場合には、予め作成してある設定値のセットから、それに対応する設定値（図2と図4で示したシーケンシャルに対応した値）をハイパバイザで選んで設定し、論理分割の処理を行う。

【0074】

具体的には、大容量アクセス向きの論理区画、小容量ファイルアクセス向きの論理区画、中間的な容量のファイルアクセスの論理区画の三つの論理区画を作成したいときには、図8のファイル制御プロセッサ、ディスク制御プロセッサの割り当て設定例で示したような、ファイル制御プロセッサの割り当てがディスク制御プロセッサをより少ない論理区画1、ファイル制御プロセッサの割り当てがディスク制御プロセッサの割り当てより多い論理区画2、同等の割り当て量の論理区画3の割り当て設定値をハイパバイザの参照する領域内（図9に後述する各論理区画への各資源の割り当て情報等を格納しておく記憶領域内）に準備しておく。

【0075】

管理者等は、割り当て設定を実際に行うに当っては、大容量ファイルアクセス向き、少容量アクセス向き、両者の中間の三つの論理区画といった、論理区画に要求する特性を簡便に指定する。するとハイパバイザがその指定に対応した割り当て設定値を自動的に選ん

で設定する。

これにより、管理者が所望の性能・特性を有する論理区画を簡単に指定することができる。

○ **【0076】**

また管理者等に割り当て設定の入力は、それぞれの論理区画が正しく動作できるだけの資源をかならず割り当てるような注意が必要となる。例えば、ファイル制御プロセッサやディスク制御プロセッサの割り当てを0にすることはできない。この点、上述のような自動設定においては、予め割り当て資源量に下限を設け、自動的にこれを守るようにしてもよい。また、図8で図示しているような入力例においては、当該NAS装置においてそれぞれの資源の割り当て量の下限を予め定義しておき、これを下回る割り当て量の場合には警告を出したり、そのような割り当て入力をチェックして受け付けないようにしてよい。

○ これにより、管理者等は安全に論理区画を設定することが出来る。

**【0077】**

図9は、上記のように管理者等が設定した論理区画への各資源の割り当てを示す情報を図示した例である。図9に示すような物理的資源と論理区画との対応関係は、ハイパバイザが管理端末から受信した情報に基づいて作成する。具体的には、それぞれのハイパバイザが、NAS装置の物理的資源の構成に関する情報を有しており、管理者等の入力情報とその構成に関する情報に基づいて物理的資源を各論理区画に割り当て、図9のような対応関係を作成する。尚、図9において、縦軸の項目はNAS装置が有する装置構成によってその項目数が増減し、横軸の論理区画は、管理者等の指定によりその数が変更される。

**【0078】**

そして、図9のような対応関係の情報は、先に説明したように、ファイル制御メモリ102、キャッシュメモリ105、ディスク群107のディスクドライブ又はその他の記憶媒体のいずれか一つ又は複数のハイパバイザが専ら用いる領域に格納される。それぞれのハイパバイザは格納された情報を参照して、各論理区画に利用させる資源を決定し、割り当ての処理を行う。

**【図面の簡単な説明】**

**【0079】**

【図1】NASの構成例を示す図である。

【図2】ファイル制御メモリとキャッシュメモリの割り当ての例を示す図である。

【図3】ファイル制御プロセッサとディスク制御プロセッサの割り当ての例を示す図である。

【図4】プロセッサ間通信接続部の割り当ての例を示す図である。

【図5】ディスクドライブ群の割り当ての例を示す図である。

【図6】ホストインターフェース部、ファイル制御プロセッサの論理分割による上位装置のユーザ認証の概念を示す図である。

【図7】NASの構成の例を示す図である。

【図8】NASの論理分割の設定入力画面の例を示す図である。

【図9】NASの各資源の論理分割情報の例を示す図である。

**【符号の説明】**

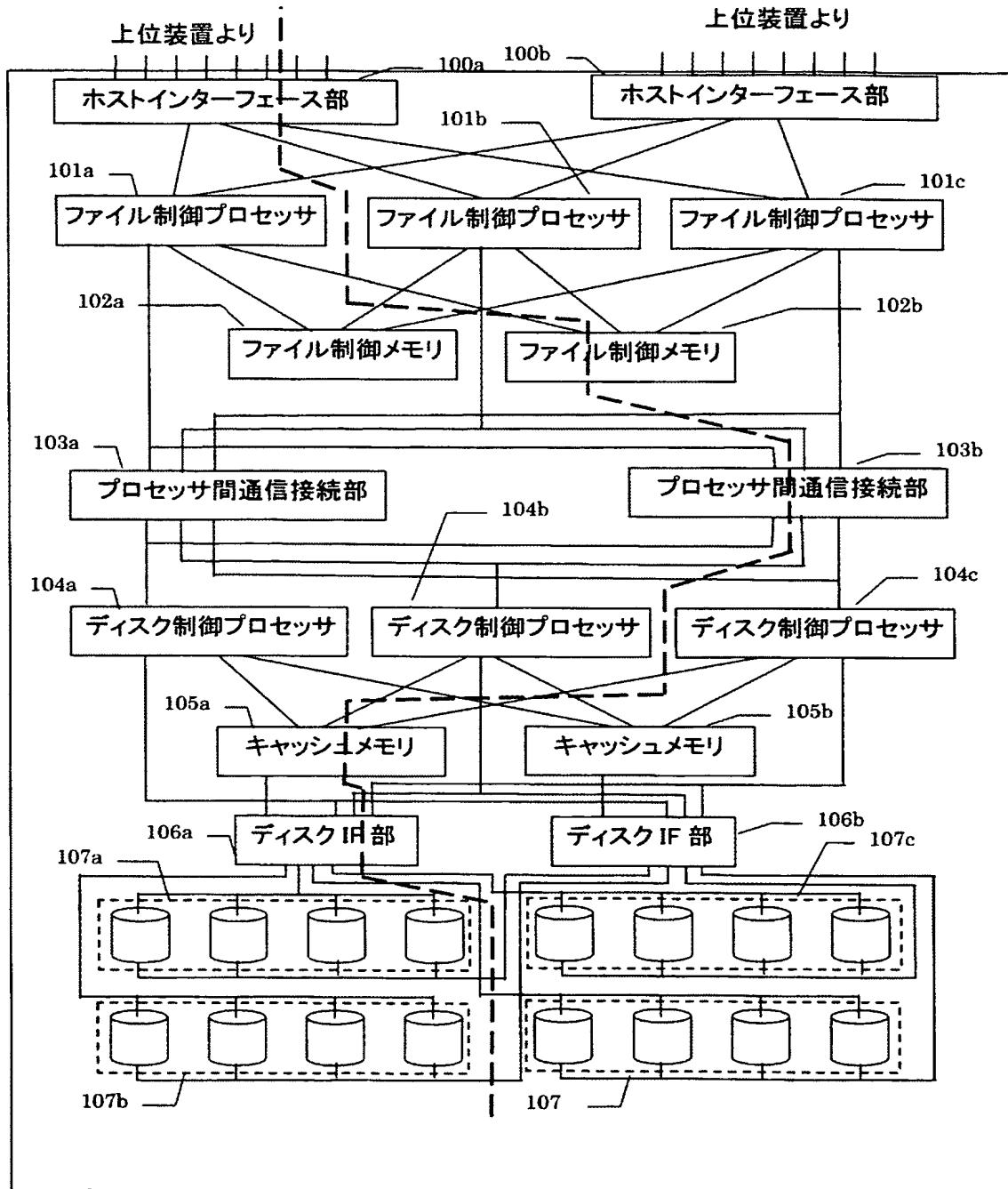
**【0080】**

100…ホストインターフェース部、101…ファイル制御プロセッサ、102…ファイル制御メモリ、103…プロセッサ間通信接続部、104…ディスク制御プロセッサ、105…キャッシュメモリ、106…ディスクIF部、107…ディスクドライブ群、901…統合制御プロセッサ、902…統合メモリ。

【書類名】 図面  
【図 1】

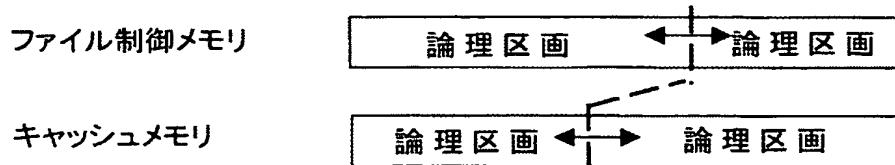
図 1

(論理分割境界)



【図 2】

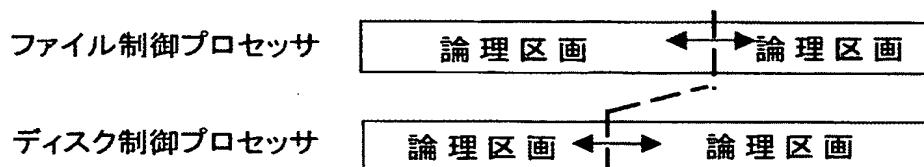
図 2



	データ読み出し 速度重視	応答速度 重視	広範囲ランダム アクセス	シーケンシャル アクセス
ファイル制御メモリ	割り当て小	割り当て大	割り当て小	割り当て大
キヤツシュメモリ	割り当て大	割り当て小	割り当て小	割り当て大

【図 3】

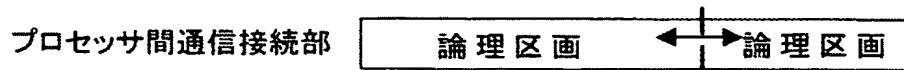
図 3



	少数の大容量 ファイルアクセス	多数の小容量 ファイルアクセス	低要求性能	高要求性能
ファイル制御プロセッサ	割り当て小	割り当て大	割り当て小	割り当て大
ディスク制御プロセッサ	割り当て大	割り当て小	割り当て小	割り当て大

【図 4】

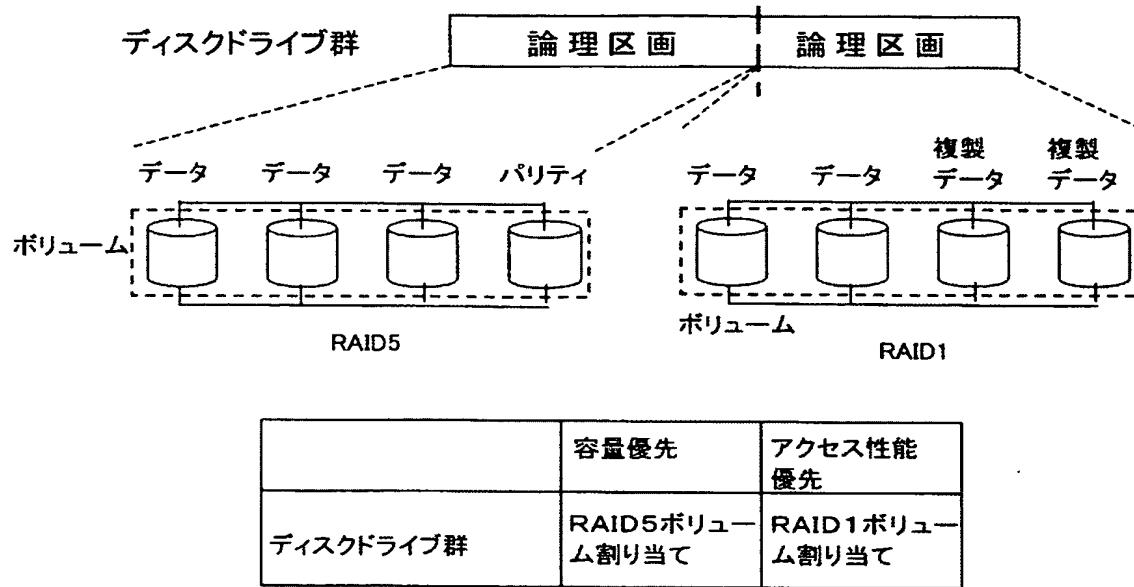
図 4



	大容量シーケン シャルアクセス	少容量シーケン シャルアクセス	ランダム アクセス
プロセッサ間通信接続部	割り当て大	割り当て中	割り当て小

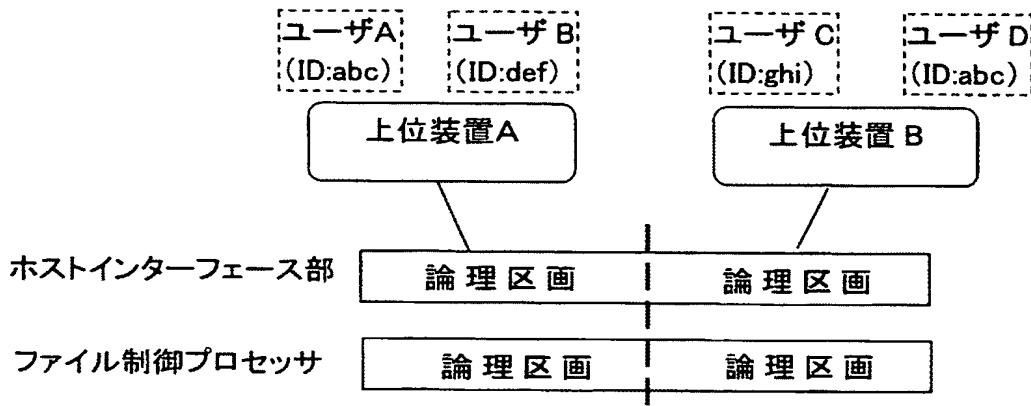
【図5】

図 5



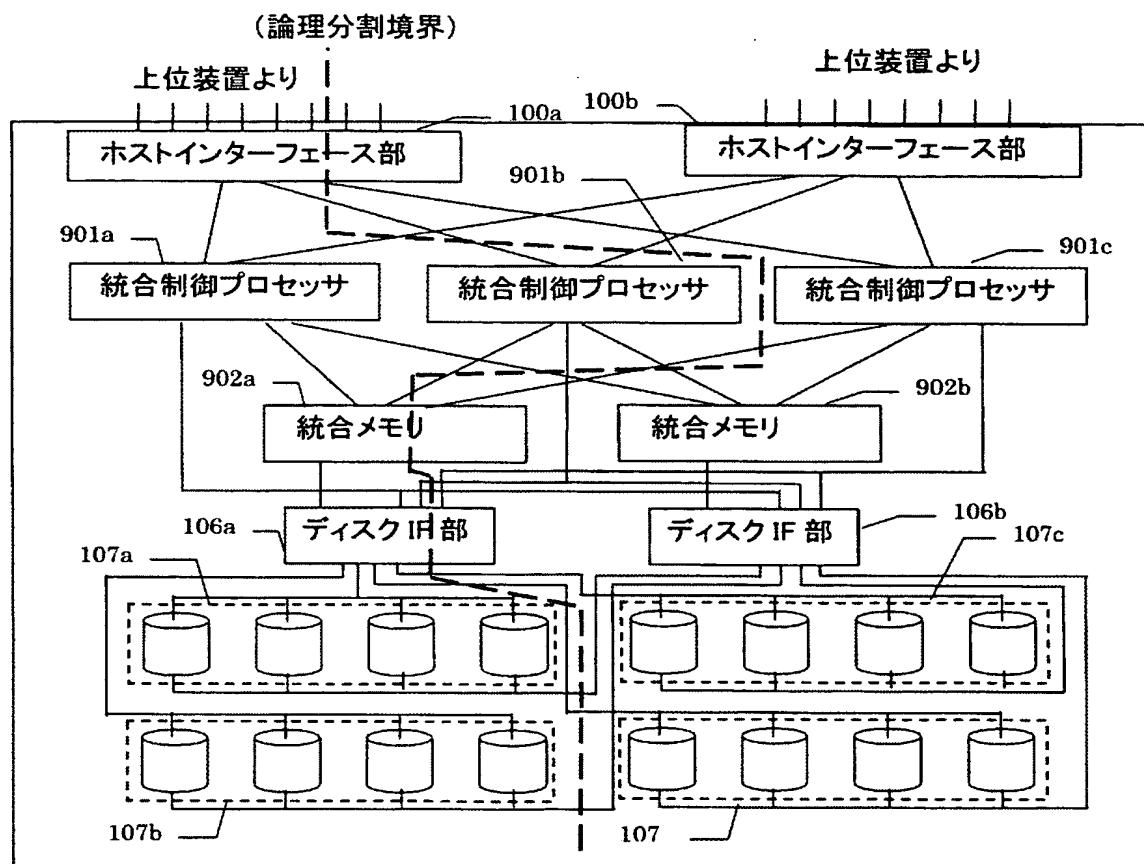
【図6】

図 6



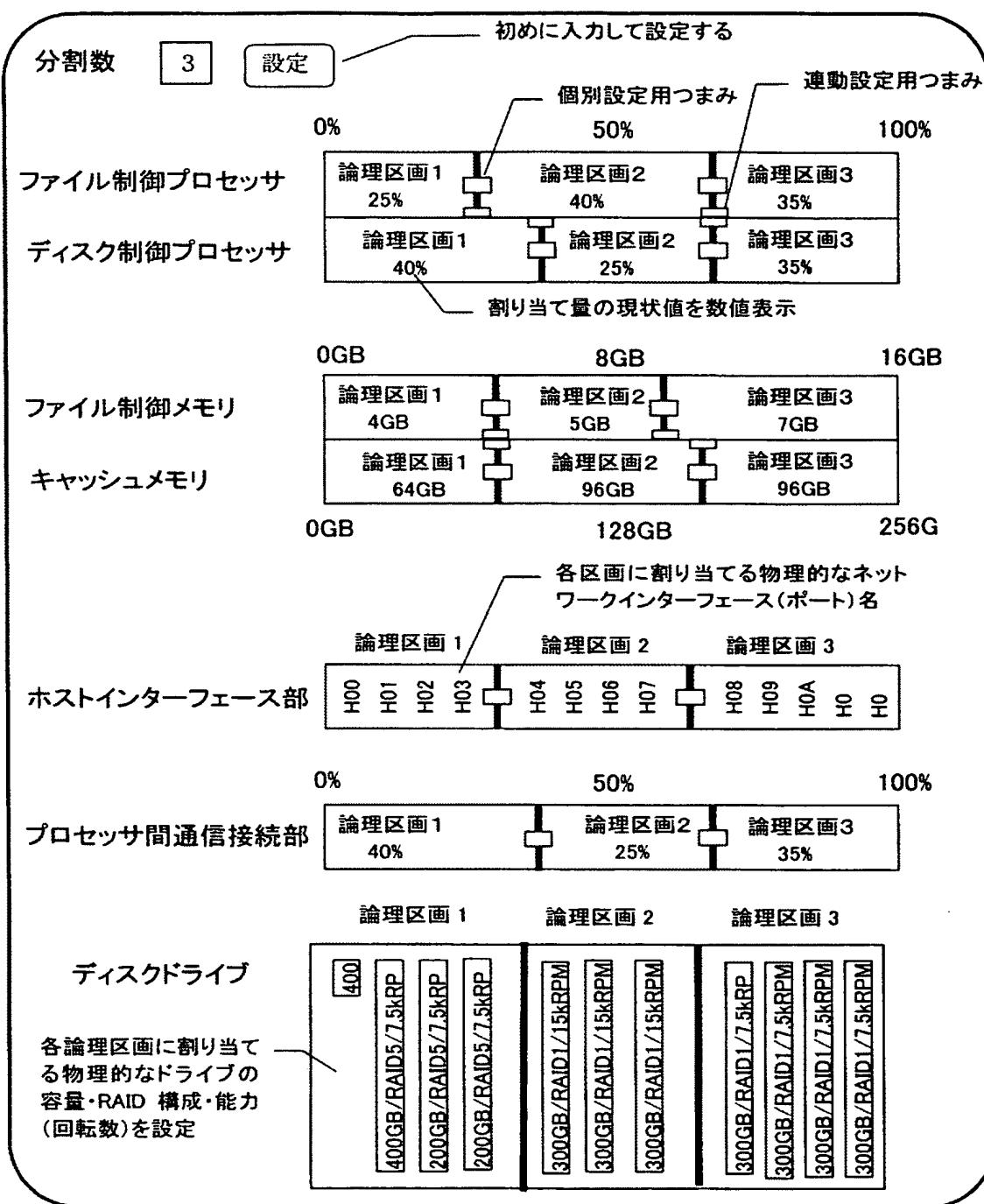
【図 7】

図 7



【図 8】

図 8



【図 9】

図 9

	論理区画1	論理区画2	論理区画3	未割り当て
ファイル制御プロセッサ1	100%	0%	0%	0%
ファイル制御プロセッサ2	0%	50%	50%	0%
ファイル制御プロセッサ3	0%	0%	0%	100%
ディスク制御プロセッサ1	100%	0%	0%	0%
ディスク制御プロセッサ2	0%	100%	0%	0%
ディスク制御プロセッサ3	0%	0%	100%	0%
ファイル制御メモリ	4GB	5GB	7GB	2GB
キャッシュメモリ	64GB	96GB	96GB	0GB
ホストインターフェース部	H00 H01 H02 H03	H04 H05 H06 H07	H08 H09 H0A H0B	H0C
プロセッサ間通信接続部	40%	25%	35%	0%
ディスクドライブ	400GB/RAID5(DR#1,2,3,4)/7.5k RPM 400GB/RAID5(DR#5,6,7,8)/7.5k RPM 200GB/RAID5(DR#9,10,11,12)/7.5k RPM 200GB/RAID5(DR#13,14,15,16)/7.5k RPM	300GB/RAID1(DR#21,22)/15k RPM 300GB/RAID1(DR#23,24)/15k RPM 300GB/RAID1(DR#25,26)/15k RPM	300GB/RAID1(DR#31,32)/7.5k RPM 300GB/RAID1(DR#33,34)/7.5k RPM 300GB/RAID1(DR#35,36)/7.5k RPM 300GB/RAID1(DR#37,38)/7.5k RPM	

【書類名】要約書

【要約】

【課題】

複数の計算機に使用されるNASにおいて、計算機相互の干渉を排除し、入出力の能力の保証・データ破壊や障害の局所化を行い、また装置内部のプロセッサ、メモリ等の資源の利用率を高める。

【解決手段】

ホストインターフェース部100、ファイル入出力要求を受け取ってデータ入出力要求に変換するファイル制御プロセッサ101、変換情報を格納するファイル制御メモリ102、ディスクドライブ群107、ディスク制御プロセッサ104、ディスクドライブ群とディスク制御プロセッサを接続するディスクIF部106、キャッシングメモリ105、プロセッサ間通信接続部103を備え、それぞれを論理的に分割することにより、仮想的な2個以上のNASとして動作させる。

【選択図】 図1

## 認定・付加情報

特許出願の番号	特願 2003-362750
受付番号	50301756105
書類名	特許願
担当官	第七担当上席 0096
作成日	平成15年10月24日

## &lt;認定情報・付加情報&gt;

【提出日】	平成15年10月23日
-------	-------------

特願 2003-362750

出願人履歴情報

識別番号 [000005108]

1. 変更年月日 1990年 8月31日

[変更理由] 新規登録

住 所 東京都千代田区神田駿河台4丁目6番地  
氏 名 株式会社日立製作所